

QUALITY-OF-SERVICE MONITOR

Field of the Invention

5 The present invention relates to voice-over-Internet-protocol telephony systems.

Background of the Invention

10 In voice-over-Internet-protocol (VoIP) telephony systems, voice calls are digitized, assembled into packets, and then directed over Internet protocol (IP) networks instead of being sent via the Switched Circuit Network (SCN). The SCN includes networks such as the Integrated Services Digital Network (ISDN) and the Public Switched Telephone Network (PSTN). Internet technologies were originally designed to handle the transport of data resilient to delays, jitter, and retransmissions, all of which can result in packet loss. Voice and other real-time media are much more sensitive to these inevitable IP network characteristics. IP instabilities can cause gaps in
15 voice transmission, clicks and other background noises, and distortions. Acceptable quality of service (QoS) is a major concern in VoIP.

 Currently, there is no system that ensures that VoIP calls are routed with acceptable quality. What is needed is a VoIP system that checks route quality and routes VoIP calls only over routes that meet a quality criterion.

20

Summary of the Invention

 The system collects values that represent a history of packet loss statistics as well as current packet loss data over a given VoIP route. The results are processed using a smoothing algorithm and are compared with user-configurable quality criteria to determine whether the route
25 is offering an acceptable quality of service. If not, a QoS Monitor automatically blocks the

unacceptable VoIP pathway so that subsequent calls are routed through an alternative, such as the SCN, for a user-defined period of time.

Brief Description of the Drawings

5 FIGURE 1 is a block diagram of a telecommunications system of the background art, including a voice-over-IP connection.

FIGURE 2 is a block diagram of a voice-over-IP system illustrating a Quality-of-Service Monitor of the present invention.

10 FIGURE 3 is an illustration of the computation of smoothed packet loss in accordance with an algorithm of the present invention.

FIGURE 4 is a flowchart of a method of the present invention.

FIGURE 5 is a flowchart of a more detailed method of the invention.

Detailed Description of the Preferred Embodiments

15 A system 100 of the background art is shown in Figure 1. System 100 picks up an analog voice signal, converts it to digital, and places it in packets. The packets are then routed to the recipient via the IP network and reconverted to an analog voice signal. The IP network uses User Datagram Protocol/Internet Protocol (UDP/IP) for transferring packets. VoIP manages the transmission of voice in packets using UDP/IP.

20 Various standards promulgated by the International Telecommunications Union (ITU) allow callers from different systems to communicate. As schematically depicted in FIGURE 1, callers using a variety of devices including analog telephones, computers, laptop computers, ISDN telephones, and wireless devices can communicate with each other over VoIP connections.

System 100 includes an ITU Standard H.323 terminal 102 connected to the switched

circuit network 104 and to the Internet 106. Terminal 102 can be, for example, a PC, a workstation, or an Ethernet-enabled telephone. According to ITU standards, H.323 terminals include at least one voice compressor/decompressor (CODEC) that sends and receives packetized audio; all H.323 terminals support Registration Admission Status Protocol (RAS), Real-Time Transport Protocol (RTP), and RTP Control Protocol (RTCP); all H.323 terminals support real-time, two-way audio communications with other H.323 entities, and may support multimedia such as audio, video, and data, or any combination of the three. An H.323 terminal typically includes a microphone/speaker including a CODEC such as ITU standard G.711 or G.729; a camera/display including a video CODEC such as ITU standard H.261; a system control user interface, including an ITU standard H.245 control channel, an ITU standard H.225 call control, and an RAS control.

A multipoint control unit (MCU) can be added to allow for conferencing functions between three or more terminals. An MCU can control and mix video, audio, and data from H.323 terminals. Typically, an MCU includes a multipoint controller that handles signaling and control for conferences, and a multipoint processor that accepts streams from LAN endpoints, replicates them and forwards them to the appropriate endpoints. In the illustrated embodiment, terminal 102 includes built-in multipoint control, implementing many of the functions of the MCU.

As those skilled in the art are aware, gatekeepers can also be added to manage nodes and to provide address translation and routing and admission control.

Gateways 108A-108D are interfaces between the H.323 system and external networks. As seen in FIGURE 1, gateway 108A interfaces with ISDN network 110, which in turn connects to an ISDN telephone 112; gateway 108B interfaces with PSTN 114, which in turn connects to an analog telephone 116; gateway 108C interfaces with an enterprise network 118, which connects to a computer 120, a digital telephone 122, and an analog telephone 124; and gateway

108D interfaces with wireless network 126, which in turn connects to a wireless communications device 128.

Gateways 108A-108D use CODECs to sample the signal from non-H.323 devices, digitize the signal, compress the data, and put the data into packets. Each gateway addresses the packets to the destination and then sends them over an IP network. The gateways also provide translation and protocol conversion between H.323 and non-H.323 devices.

For minor packet loss, the CODEC can perform some compensating functions, such as retransmission of the previous packet or interpolation. However, when packet loss exceeds a threshold (typically about 5%), or when a set of sequential packets is lost, voice quality can be significantly degraded.

A system 200 incorporating the present invention is illustrated in FIGURE 2. In the example depicted in FIGURE 2, an H.323 terminal 202 is in communication with an analog telephone 204. Terminal 202 connects to the PSTN 206 and an IP network 208. A gateway 210, including a processor 212, interfaces between PSTN 206 and the IP network 208. A processor can also be incorporated in terminal 202.

System 200 evaluates packet loss data using a smoothing algorithm to assess system reliability and by extension, voice quality. The smoothing algorithm provides a way to adjust for transitory effects that may skew the results. Before blocking call transmission, it is advantageous to know that the connection is consistently bad, not just experiencing an aberration. For instance, if a call has dropped 1 percent of its packets during each of 100 RTCP intervals but drops 20 percent of its packets during the next interval, the smoothing algorithm will handle this as an irregularity. Instead of automatically increasing the average packet loss to a value which does not accurately reflect the very stable history of this connection, the occurrence of the packet loss spike will only affect the average percent packet loss to a minimal degree.

In a preferred embodiment, the system uses the Sliding Window Algorithm to set each output to a weighted average of the previous output and the latest sample (S_i). In the system's implementation of the Sliding Window Algorithm, a new computed smoothed loss V_i is set to the weighted average of V_{i-1} and S_i according to the following formula, known as the Sliding Window

5 Exponential Average Algorithm:

$$V_i = \alpha V_{i-1} + (1-\alpha)S_i$$

in which the variables are defined as follows:

V_i is the computed smoothed percent packet loss.

10 V_{i-1} is the smoothed percent packet loss calculated during the previous RTCP interval. This value represents the quality of the call to the remote IP address from call initiation up to the current RTCP interval.

S_i is the percent packet loss representing the current interval.

α is the Route Blocking Sensitivity.

15 Several user-configurable variables are used in the implementation of the Sliding Window Exponential Average Algorithm: (1) the Blocking Duration, which is the time in minutes that a route will be blocked if its quality is deemed unacceptable; (2) the Route Blocking Threshold, which represents the packet loss percent threshold at which the route will be blocked; (3) the Route Blocking Sensitivity (" α " above), which determines how fast the smoothing algorithm will

20 react to changes in the rate of packet loss, thereby establishing the pervasiveness of the "memory" of the connection's packet loss history (the Route Blocking Sensitivity is limited to values between 0 and 1); (4) the Route Blocking Clamp, which affects how the smoothing algorithm reacts to sudden spikes in the packet loss, setting a maximum percent packet loss that will be used

in the smoothing algorithm for a given interval. If the actual packet loss for an interval is greater than the value of the Route Blocking Clamp, the packet loss percentage value will be set to the clamped value. The Route Blocking Clamp exists to avoid distortions of the smoothing algorithm's results due to an unusually high packet loss percentage during a single interval; (5) the Minimum Call Duration, the interval for which a call must be maintained before it may be declared unacceptable. It is expressed in seconds, which are converted internally to a number of intervals; and (6) the RTCP Interval, which is the Real-time Transport Control Protocol reporting interval, generally between 5 and 9 seconds. The RTCP generates statistics relating to each interval, including the number of packets expected, the number of packets lost, and the average jitter value.

An example of the computation of smoothed percent packet loss in the Sliding Window Exponential Algorithm is illustrated in FIGURE 3. A table 302 illustrates how the smoothed packet loss values are calculated. There are four samples S_i : 1, 2, 3, 1. Smoothed packet loss values V_i are calculated from these S_i values. Note the first sample S_0 is used as the "smoothed" V_0 without applying the formula as there is no history yet. The column headed "Expansion" illustrates that the effective coefficient in V_i of a previous sample S_{i-k} ($0 \leq k < i$) is $(1-\alpha) \cdot \alpha^k$ (thus the name *exponential* average). For example, in FIGURE 3, where $\alpha = \frac{2}{3}$, the effective coefficient in V_3 of S_1 is $\frac{1}{3}(\frac{2}{3})^2$. A graph 304 illustrates the smoothing function. (In a preferred embodiment, $\alpha = \frac{1}{2}$; FIGURE 3 uses $\alpha = \frac{2}{3}$ to allow α and $1-\alpha$ to be distinguished.)

So long as the algorithm yields acceptable values, calls continue to be routed over the IP network. If, on the other hand, the value exceeds a threshold, a QoS Monitor blocks routing over the IP network and routes calls over an alternative network, such as the SCN or a wireless network. The QoS Monitor is a computer program that can be implemented in H.323 terminal

202, or in gateway 210, or distributed across 202 and 210. (In an H.323-to-H.323 call, gateways may not be needed; alternatively, an analog-to-analog call may be a VoIP call if both devices are mediated by a gateway.) Where an alternative packet-based routing system exists, the system may route calls over alternative packet-based routes.

5 When the QoS Monitor blocks network routing, a timer included in gateway 210 or, alternatively, in the terminal is started. When the timer runs out, the system again allows routing through the IP network. Packet loss continues to be monitored, and if packet loss again reaches unacceptable levels according to the Sliding Window Exponential Average Algorithm, the QoS Monitor again blocks network routing.

10 A method 400 of the invention is shown in Figure 4. At step 402, parameters are initialized. At step 404, a voice-over-Internet protocol connection is established. At step 406, the packet percentage loss is checked; in a preferred embodiment, the RTCP gathers the packet loss data and the QoS Monitor evaluates it. At step 408, the processor updates the data set.

15 At step 410, the processor associated with the QoS Monitor compares the updated data set to a selected threshold value. In a preferred embodiment, packet loss data are represented by a value computed by application of a smoothing algorithm such as the Sliding Window Algorithm. Values can also be determined by other smoothing algorithms.

20 At a step 412, the processor associated with the QoS Monitor determines whether the quality of the updated set is acceptable. If yes, calls continue to be routed over the IP network, at step 414, and the method loops back to step 406, and continues to monitor the packet percentage loss.

 If the quality is not acceptable, the method initializes a timer, at step 416. The QoS Monitor blocks the unacceptable IP network route until the timer times out, at step 418, so that future calls are not carried over it. Future calls can alternatively be switched to the PSTN, or can

be routed over an alternative IP network route.

A more detailed method 500 of the invention is shown in FIGURE 5. The method starts at step 502. The method waits for a connection request at step 504. At step 506, the connection is requested. At step 508, the method checks whether the QoS Monitor is currently blocking routes over the IP network.

If yes, the method routes the call over the PSTN or alternative, at step 510. If IP network routing is not being blocked by the QoS Monitor, parameters are initialized, at step 512. In a preferred embodiment, the parameters are Minimum Call Duration (MCD); RTCP Interval; Blocking Duration; Route Blocking Clamp; Sensitivity (α); and Threshold.

At step 514, a voice-over-Internet-protocol (VoIP) connection is established, and the interval is initialized to 0. At step 516, the method waits for an RTCP interval. At step 518, the RTCP determines the percentage packet loss, L_i . At a step 520, the method checks whether the value for L_i is less than the value for the Route Blocking Clamp. If the value for L_i is less than the value for the Route Blocking Clamp, the value for S_i is set to L_i , at step 522. Otherwise, S_i is set to the value of the Route Blocking Clamp, at step 524. In either case, the method moves to step 526, where, if it is the initial interval ($i=0$), V_{i-1} is set to the value of S_i .

The method then proceeds to step 528, where the new computed smoothed loss V_i is set to the weighted average of V_{i-1} and S_i according to the Sliding Window Exponential Average Algorithm:

$$V_i = \alpha V_{i-1} + (1-\alpha)S_i$$

The method then proceeds to step 530, where it is determined whether blocking by the QoS Monitor is on. If yes, the method checks whether the current call is over, at step 532. If the current call is over, the method returns to step 504, and waits for a connection request. If, on the

other hand, the current call is not over, the method proceeds to step 534, where interval number i is set to $i + 1$, and then continues to gather statistics about the quality of the connection. The method then returns to step 516, and waits for another RTCP interval.

If, at step 530, the blocking is not on, the method proceeds to step 536, where it is
 5 determined whether V_i is greater than the Route Blocking Threshold. If V_i is not greater than the Route Blocking Threshold, the method proceeds to step 532 and proceeds as described above.

If V_i is greater than the Route Blocking Threshold, the method proceeds to step 538, and checks whether the call time is greater than or equal to the Minimum Call Duration (MCD). If not, the method again proceeds to step 532. If the call time is greater than or equal to the MCD,
 10 the method turns QoS Monitor blocking on for the blocking duration, at step 540. After the blocking is turned on at step 540, the method proceeds to step 532 and the steps following.

The configuration of components is a matter of design choice; other configurations will be well known to those skilled in the art. The apparatus and method are compatible with other protocols and devices. Other smoothing algorithms can be used. The system can be used to
 15 assess the quality of calls in progress, and existing calls can be dynamically monitored and rerouted. The apparatus and method can be used with IP networks including intranets and the Internet. The apparatus and method can simultaneously monitor multiple calls, *e.g.*, to a common IP destination, to improve the sampling statistics. If separate values of smoothed packet loss V_i are tracked for each call, then the value consulted for prospective calls can be either the most
 20 recently written V_i or, alternatively, an average of V_i values across current calls. Alternatively, a single V_i can be kept, pooling data for the different calls. Other network systems are compatible with the invention. The compatibility of the invention with the use of other components such as routers and gatekeepers will be known to those skilled in the art. Those skilled in the art will be

[illegible]